Sound to Sense, Sense to Sound

Sound to Sense, Sense to Sound
A State of the Art in Sound and Music Computing

*** Draft ***

February 2008

Pietro Polotti and Davide Rocchesso, editors

Information Society Technologies

Chapter 2

Learning music: prospects about implicit knowledge in music, new technologies and music education

Emmanuel Bigand & Philippe Lalitte
LEAD CNRS, Pôle AAFE-Esplanade Erasme, Université de Bourgogne

Barbara Tillmann
CNRS-UMR 5020 & IFR 19, Lyon, France

About this chapter

The chapter proposes an overview of musical learning by underlining the force of the cognitive system, which is able to learn and to treat complex information at an implicit level. The first part summarises recent research in cognitive sciences, that studies the processes of implicit learning in music perception. These studies show that the abilities of non-musicians in perceiving music are very often comparable to those of musicians. The second part illustrates by means of some examples the use of multimedia tools for learning tonal and atonal music; these tools take advantage of the interaction between visual and auditive modalities.

Introduction

Delineating the musical abilities that are specifically linked to an intensive and formal training from those that emerge through mere exposure to music is a key issue for music cognition, music education, and all of the disciplines involved in sound and music computing, particularly for disciplines which deal with content processing of audio signals (see Chapter 3), machine learning (see Chapter 4), and sound design (see Chapter 10). Non-musicians do not learn a formal system with which they can describe and think about musical structures. Nevertheless, they have a considerable amount of experience with music: they hear music every day of their lives. They have all sung as children and in school, they are moving and dancing to musical rhythms, and most of them have attended concerts. Nowadays the new wearable digital audio players make it easy to listen to a large amount of music in all circumstances. How sophisticated are the emerging abilities to process music that result from this exposure when compared to the abilities caused by an intensive formal musical training? Given the huge differences in training, finding disparities between musically trained and untrained listeners would not be really surprising. However, research in auditory cognition domain has shown that even non-musician listeners have knowledge about the Western tonal musical system. Acquired by mere exposure, this implicit knowledge guides and shapes music perception. This chapter presents recent research studying implicit learning in music, and some examples of multimedia tools for learning Western tonal music as well as contemporary music. These tools are based on advances in cognitive psychology concerning the acquisition and the representation of knowledge, and the role of memory and of attention processes.

2.1 Implicit processing of musical structures

2.1.1 How do non-musician listeners acquire implicit knowledge of music?

Implicit learning processes enable the acquisition of highly complex information without complete verbalisable knowledge of what has been learned (Seger, 1994). Two examples of highly structured systems in our environment are language and music. Listeners become sensitive to the underlying regularities just by mere exposure to linguistic and musical material in everyday life. The implicitly acquired knowledge influences perception and interaction with the environment. This capacity of the cognitive system has been studied in the laboratory with artificial material containing statistical structures, such as finite-state grammars or artificial languages (i.e. Altmann et al., 1995; Reber, 1967, 1989; Saffran et al., 1996). Tonal acculturation is one example of the cognitive capacity to become sensitive to regularities in the environment. Francès (1958) was one of the first underlining the importance of statistical regularities in music for tonal acculturation, suggesting that mere exposure to musical pieces is sufficient to acquire tonal knowledge, even if it remains at an implicit level. In the music cognition domain, numerous research has provided evidence for non-musicians' knowledge about the tonal system (see Bigand and Poulin-Charronnat, 2006, for a review).

2.1.2 Implicit learning of Western pitch regularities

Western tonal music constitutes a constrained system of regularities (i.e. frequency of occurrence and co-occurrence of musical events, and psychoacoustic regularities) based on a limited number of elements. This section presents the tonal system from the perspective of cognitive psychology: it underlines the basic regularities between musical events, which appear in most musical styles of everyday life (e.g. classical music, pop music, jazz music, Latin music, etc.) and which can be acquired by implicit learning processes. The Western tonal system is based on 12 pitches repeated cyclically over octaves. Strong regularities of co-occurrence and frequencies of occurrence exist among these 12 pitch classes (referred to as the tones C, C#/Db, D, D#/Eb, E, F, F#/Gb, G, G#/Ab, A, A#/Bb, B): tones are combined into chords and into keys, forming a three-level organisational system (Figure 2.1). Based on tones and chords, keys (tonalities) have more or less close harmonic relations to each other. Keys sharing numerous tones and chords are said to be harmonically related. The strength of harmonic relations depends on the number of shared events. In music theory, major keys are conceived spatially as a circle (i.e. the circle of fifths), with harmonic distance represented by the number of steps on the circle. Inter-key distances are also defined between major and minor keys. The three levels of musical units (i.e. tones, chords, keys) occur with strong regularities of co-occurrence. Tones and chords belonging to the same key are more likely to co-occur in a musical piece than tones and chords belonging to different keys. Changes between keys are more likely to occur between closely related keys (e.g. C and G major) than between less-related ones (e.g. C and E major). Within each key, tones and chords have different tonal functions creating tonal and harmonic hierarchies. These within-key hierarchies are strongly correlated with the frequency of occurrence of tones and chords in Western musical pieces. Tones and chords used with higher frequency (and longer duration) correspond to events that are defined by music theory as having more important functions in a given key (Budge, 1943;

Francès, 1958; Krumhansl, 1990).

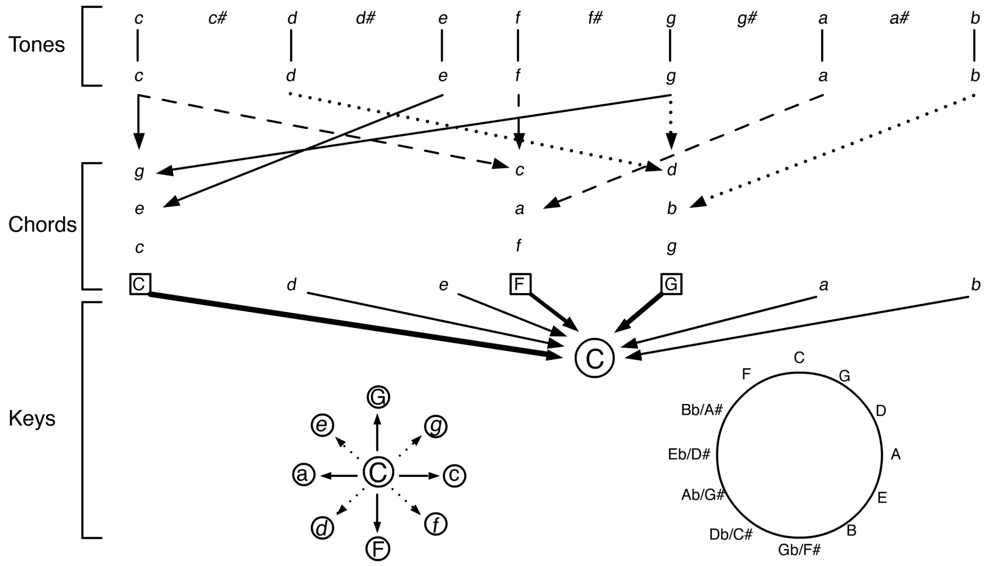

Figure 2.1: Schematic representations of the three organisational levels of the tonal system. a) 12 pitch classes, followed by the diatonic scale in C Major. b) construction of three major chords, followed by the chord set in the key of C Major. c) relations of the C Major key with close major and minor keys (left) and with all major keys forming the circle of fifths (right). Tones are represented in italics, minor and major chords/keys in lower and upper case respectively (from Tillmann et al., 2001).

This short description reveals a fundamental characteristic of the Western tonal music: functions of tones and chords depend on the established key. The same event can define an in-key or an out-of-key event and can take different levels of functional importance. For listeners, understanding context dependency of musical events' functions is crucial for the understanding of musical structures. Music cognition research suggests that mere exposure to Western musical pieces suffices to develop implicit, but nevertheless sophisticated, knowledge of the tonal system. Just by listening to music in everyday life, listeners become sensitive to the regularities of the tonal system without being necessarily able to verbalise them (Dowling and Harwood, 1986; Francès, 1958; Krumhansl, 1990). The seminal work by Krumhansl, Bharucha and colleagues has investigated the perception of relations between tones and between chords as well as the influence of a changing tonal context on the perceived relations (see Krumhansl, 1990, for a review). The data showed the cognitive reality of tonal and harmonic hierarchies for listeners and the context dependency of musical tones and chords in perception and memorisation.

2.1.3 Connectionist model of musical knowledge representation and its acquisition

Bharucha (1987) proposed a connectionist account of tonal knowledge representation. In the MUSACT model (i.e. MUSical ACTivation), tonal knowledge is conceived as a network of interconnected units (Figure 2.2). The units are organised in three layers corresponding to tones, chords, and keys. Each tone unit is connected to the chords of which that tone is a component. Analogously, each chord unit is connected to the keys of which it is a member. Musical relations emerge from the activation that reverberates via connected links between tone, chord and key units. When a chord is played, the units representing the sounded component tones are activated and activation reverberates between the layers until equilibrium is reached (see Bharucha, 1987; Bigand et al., 1999, for more details). The emerging activation patterns reflect tonal and harmonic hierarchies of the established key: for example, units representing harmonically related chords are activated more strongly than units representing unrelated chords. The context dependency of musical events in the tonal system is thus not stored explicitly for each of the different keys, but emerges from activation spreading through the network.

In Tillmann et al. (2000), we take advantage of the learning possibilities of artificial neural networks (e.g. connectionist models) to simulate tonal knowledge acquisition in non-musician listeners. For this purpose, unsupervised learning algorithms seem to be well suited: they extract statistical regularities via passive exposure and encode events that often occur together (Grossberg, 1970, 1976; Kohonen, 1995; Rumelhart and Zipser, 1985; von der Malsberg, 1973). Self-organizing maps (Kohonen, 1995) are one version of unsupervised learning algorithms that leads to a topological organisation of the learned information.

To simulate tonal acculturation, a hierarchical network composed of two self-organizing maps was exposed to short musical sequences (i.e., chord sequences). After learning, the connections in the network have changed and the units have specialised for the detection of chords and keys (the input layer coded the tones of the input material). The learned architecture is associated with a spreading activation process (as in MUSACT) to simulate top-down influences on the activation patterns. Interestingly, the learned connections and the activation patterns after reverberation mirror the outcome of the hardwired network MUSACT, which has been conceived as an idealised end-state of implicit learning processes (see Tillmann et al., 2000). In collabora-

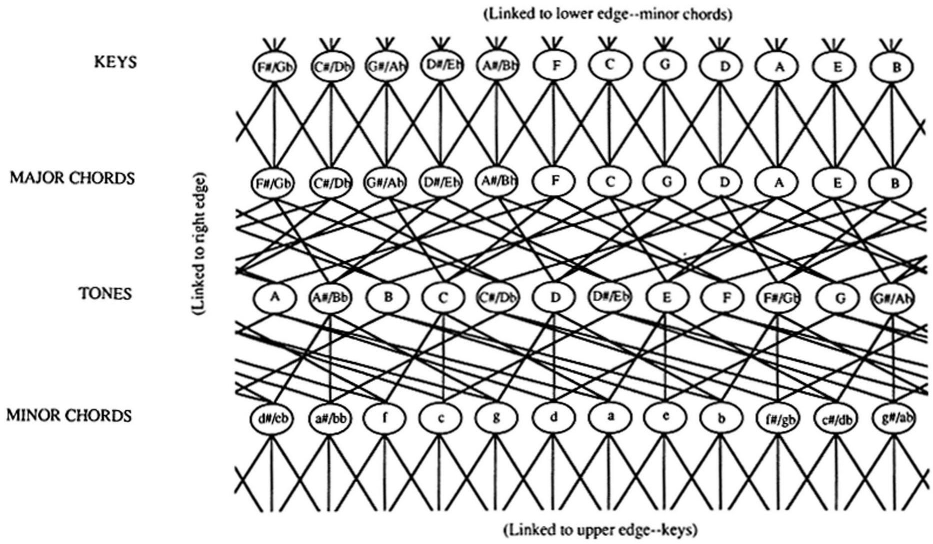

Figure 2.2: MUSACT model of tonal knowledge activation. The tone layer is the input layer, which is connected to the chord layer (consisting of major and minor chords). The chord layer is connected to the key layer (the third layer). Adapted from Bharucha (1987).

tion with Michel Paindavoine (LE2I-CNRS, Dijon) and Charles Delbé (LEAD-CNRS, Dijon), the authors are currently working on several extensions of this connectionist approach. One of the projects concerns the construction of a set of "artificial musical ears" for this modelling approach. A step of auditory pre-processing will allow us to decode sound files and to work with musical stimuli having greater acoustic complexity. On the basis of this richer input, a network will be trained with a corpus of real recordings containing a variety of musical pieces.

2.1.4 Studying implicit learning processes with artificial materials

Implicit learning processes are supposed to be at the origin of listeners' tonal knowledge, acquired in everyday life. Implicit learning processes are studied more closely in the laboratory with artificial materials containing statistical regularities. In the seminal work by Reber (1967), participants were asked to memorise grammatical letter strings in a first phase of the experiment. They were unaware that rules existed. During the second phase of the experiment, they were informed that the previously seen sequences had been produced by a rule system (which was not described) and were asked to judge the grammaticality of new letter strings. Participants differentiated grammatical letter strings from new ungrammatical ones at better than chance level. Most of them were unable to explain the rules underlying the grammar in free verbal reports (e.g. Altmann et al., 1995; Dienes et al., 1991; Dienes and Longuet-Higgins, 2004; Reber, 1967, 1989).

Various findings are convergent in demonstrating the cognitive capacity to learn complex structures and regularities. The acquisition of regularities in the experimental material is not restricted to visual events (e.g. letters, lights, shapes), but has been extended to auditory events, such as sine waves (Altmann et al., 1995), musical timbres (e.g. gong, trumpet, piano, violin, voice in Bigand et al., 1998) or environmental sounds (e.g. drill, clap, steam in Howard and Ballas, 1980, 1982). Recent studies have started to consider the acoustical characteristics of the sound, such as prosodic cues (Johnson and Jusczyk, 2001; Thiessen and Saffran, 2003; Saffran et al., 1996) or acoustical similarities (Tillmann and McAdams, 2004). The aim of the studies was to test whether the relation between the statistical regularities and regularities inherent in the acoustical material could influence learning: conflicting information might hinder statistical learning, while converging information might facilitate learning. Tonal acculturation might represent a beneficial configuration: musical events appearing frequently together are also linked acoustically since they share (real and virtual) harmonics. To investigate whether convergence with acoustical features represent a facilitating or even necessary condition for statistical learning, Tillmann and McAdams (2004) systematically manipulated acoustical similarities between timbres so that they either underline the statistical regularities of the timbre units, contradict these regularities or are neutral to them. The outcome showed that listeners learned the statistical regularities of the complex auditory material and that the manipulated surface characteristics did not affect this statistical learning. The surface characteristics only affected grouping and overall preference bias for the different materials. This outcome suggests that tonal acculturation does not necessarily need the convergence between statistical and acoustical regularities. Supporting evidence can be found in acculturation to Arabic music, which is lacking the convergence between statistical and acoustic features (Ayari and McAdams, 2003). Together with the implicit learning study on twelve-tone music (see below), the data emits the rather encouraging hypothesis about the possibility to learn regularities of new musical styles.

2.1.5 Implicit learning of new musical systems

Music is an interesting medium to investigate implicit learning processes for several reasons. It is a highly complex structure of our environment that is too complex to be apprehended through explicit thoughts and deductive reasoning. Musical events *per se* are of no importance, yet musical pieces are more than a pleasing succession of coloured sounds. The psychological effects of musical sounds come from the complex multilevel relationships between musical events involved in a given piece (Meyer, 1956; Lerdahl and Jackendoff, 1983). The abstract associative and architectonic relationships that are not close in time define relevant structures in music. These relations cannot be easily articulated in an explicit way. Despite an eminent tradition in music history, as well as in contemporary music theory, to formalise the relevant structure of Western music (see Lerdahl and Jackendoff, 1983; Lerdahl, 2001; Narmour, 1990), none of these frameworks provides a complete and satisfactory account of the Western musical grammars. A further interesting feature of music for research on implicit learning is that musical structures are not always conceived for being explicitly processed. It is even of crucial importance for composers that listeners are sensitive to the structures that underlie a musical piece while still being unaware of them. And in fact, the most common impression among a general audience is that of being unable to verbally describe what they perceive. In some instances, people are even convinced that they do not perceive any underlying structure. The fact that musical events do not refer to any specific object in the external world probably contributes to the difficulty of apprehending musical structures in an explicit way.

A final interesting feature is that musical systems constantly evolve towards new musical grammars. Being faced with masterpieces that derive from an entirely new musical system is not an artificial situation for contemporary listeners and this raises a challenging issue for implicit learning theories. The considerable and persistent confusion reported by listeners to contemporary music suggests that some musical grammars may be too artificial to be internalised through passive exposure (Lerdahl, 1989). As a consequence, several cognitive constraints have been delineated, and musical grammars should obey these constraints in order to be learnable (Lerdahl, 1988, 2001). Contemporary music challenges the ability of the human brain to internalise all types of regularities. This raises a question with implications for cognitive science, music cognition, and contemporary music research.

To the best of our knowledge, very little research has directly addressed implicit learning with musical material (Bigand et al., 1998; Dienes et al., 1991). Numerous research in music cognition, however, deals indirectly with implicit learning processes by showing that explicit learning is not necessary for the development of a sensitivity to the underlying rules of Western music[1] (see section above). Only a few studies have addressed the implicit learning of new musical systems. Most of them have focused on the learning of serial music, a system that appeared in the West in the first half of the 20th century. During this period, the tonal system was overtaken by the serial system developed, in particular, by Schoenberg (Griffiths, 1978). Serial works of music obey compositional rules that differ from those that govern tonal music.

A serial musical piece is based on a specific ordering of the chromatic scale called the twelve-tone row. A twelve-tone row is an arrangement, into a certain order, of the twelve tones of the chromatic scale regardless of register (Figure 2.3). The tones of the row must be used in their chosen order (repetition of tones is allowed in certain circumstances and two or more successive tones of the row may appear as a chord), and once all twelve tones of the row have appeared, the row is repeated again and again until the end of the composition. The row may appear in any of its four basic forms: the original row, the inverted form (in which ascending intervals of the original row are replaced by equivalent descending ones and vice versa), the retrograde form (in which the tones of the original row are read backwards), and the retrograde inversion (in which the tones of the inverted form are read backwards), and each of the four forms of the row may be transposed to any of the twelve tones of the chromatic scale, thus making available forty-eight permissible patterns of one row. In theory, each tone of the row should have roughly the same frequency of occurrence over the entire piece.

Each serial composition results from a complex combination of all of these transformations which are applied to one specific tone row. Schoenberg argued that these manipulations would produce an interesting balance between perceptual variety and unity. A critical point on which he insisted was that the initial row must remain unchanged throughout the entire piece. In other words, Schoenberg's cognitive intuition was that the perceptual coherence deriving from the serial grammar was unlikely to be immediately perceived but would result from a familiarisation with the row.

Several experimental studies have addressed the psychological reality of the organisation resulting from serial musical grammar. The oldest, by Francès (1958, exp. 6), consisted of presenting participants with 28 musical pieces based on a specific tone row and requiring participants to detect four pieces that violated the row. These odd pieces were actually derived from another

[1] The tonal system designates the most usual style of music in the West, including, Baroque (Bach), Classic (Mozart) and Romantic (Chopin) music, as well as to a certain extent folk music such as pop-music, jazz and Latin-music.

row (the foil row). The analysis of accurate responses revealed that participants had considerable difficulty in detecting the four musical pieces that violated the initial row. Moreover, the fact that music theorists specialised in serial music did not respond differently from musically untrained participants suggests that extensive exposure to serial works is not sufficient for the internalisation of this new musical system. Although Francès' research is remarkable as pioneer work in this domain, the study contained several weaknesses relative to the experimental design as well as to the analysis of the data and this detracts from the impact of his conclusion. The most noticeable problem concerns the foil row, notably because it was strongly related to the tested row.

Empirical evidence supporting the perceptual reality of the rules of serial music was reported by Dowling (1972) with short melodies of 5 tones. In Dowling's experiment, participants were trained to identify reversed, retrograde and retrograde-inversion of standard melodies of 5 tones with equal duration. The melodies were deliberately made with small pitch intervals in order to improve performance. Dowling observed that musically untrained participants managed to identify above chance the rules of the serial music, with highest accuracy for the reversed transformation and the lowest for the retrograde inversion. Given that Dowling's musical stimuli were extremely short and simple, it is difficult to conclude that the rules of serial music may be internalised from a passive hearing of serial music. Moreover, in a very similar experiment using 12 tones instead of 5, Delannoy (1972) reported that participants did not succeed above chance in distinguishing permitted transformations of a standard musical sequence from those that violated the serial rules.

More recently, Dienes and Longuet-Higgins (2004) have attempted to train participants in the grammar of serial music by presenting them with 50 musical sequences that illustrated one of the transformation rules of serial music. The second half of the row was a transformation of the first half (i.e., a reverse, a retrograde or a retrograde inversion transformation). After this familiarisation phase, participants were presented with a new set of 50 sequences, some of them violating the rules of serial music (i.e., the last 6 notes were not a permitted transformation of the first 6). Participants were required to differentiate grammatical pieces (according to serial rules) from non-grammatical ones. Accuracy rates generally did not differ from chance level, which is consistent with Francès (1958)' and Delannoy (1972)'s findings.

A critical feature of the experiment of Dienes and Longuet-Higgins (2004) is that participants had never been exposed to a single tone row. Participants were trained with the transformational rules of serial music, but these rules were always instantiated with a new set of tones.

The temporal order of the first 6 notes was chosen at random. As a consequence, the referential row was constantly moving from one trial to the other. This procedure is very demanding since it consists in requiring participants to learn abstract rules which are illustrated by a constantly changing alphabet. To the best of our knowledge, there is no evidence in the implicit learning domain to show that learning can occur in this kind of situation. If participants do not have the opportunity to be exposed to an invariant tone row in the training phase, it is not surprising that they fail to exhibit sensitivity to the serial grammar in the test phase. It should be noticed that this situation violates the basic principle of serial music postulating that only one row should be used for one piece. Krumhansl and Sandell (1987) have provided the strongest support for the psychological relevance of serial rules. As illustrated in their study, experiments 1 and 2 were run with simple forms of two tone rows. That is to say, the tone rows were played with isochronous tones that never exceeded the pitch range of one octave. Experiments 3 and 4 were run with excerpts of *Wind Quintet* op. 26 and *String Quartet* op. 37 by Schoenberg. The results of the classification tasks used in Experiments 2 and 3 demonstrated that participants discriminated, above chance level, between inversion, retrograde, and retrograde inversion of the two tone rows with correct responses varying from 73% to 85% in Experiment 2, and from 60% to 80% in Experiment 3. At first glance, this high accuracy is surprising. However, it should be noticed that participants were exposed to very simple forms of the tone rows a great number of times during Experiment 1. It seems likely that this exposure helps to explain the good performance. In other words, the peculiar importance of this study lies in the suggestion that previous exposure to a tone row can be a critical feature for the perception of the rules of serial music. The question remains, however, about the type of learning that actually occurred during this prior exposure. Given that all participants had formal instruction in music education, we cannot rule out the possibility that they used their explicit knowledge of musical notation to mentally represent the structures of the two rows.

In order to define the nature (implicit/explicit) of the knowledge in learning serial music rules, Bigand, D'Adamo and Poulin (in revision) have tested the ability of musically untrained and trained listeners to internalise serial music rules with 80 two-voice pieces, especially designed by the composer D. A. D'Adamo. A set of 40 pieces defined various instantiations (transpositions) of one twelve-tone row (grammatical pieces). The other set of 40 pieces were derived from another twelve-tone row (ungrammatical pieces). As it is shown in Figure 2.3, each ungrammatical piece was matched to a grammatical piece according to their superficial features (rhythm, pitch ranges, overall form of melodic contour, duration, dynamics). The ungrammatical pieces differed from the grammatical pieces only in the twelve-tone row used. In the learning phase, 20 pieces were presented twice to the participants, who had simply to indicate whether a given piece was heard for the first time or for the second time. In the test phase, 20 pairs of pieces were presented to the participants. Each pair contained a grammatical piece which had not been heard in the training phase and a matched ungrammatical piece (Figure 2.3). Since the pieces of a pair shared the same musical surface (i.e. same pitch range, melodic contour and rhythm), even if they were derived from two different twelve-tone rows, they sounded very close. The participants were asked to indicate which piece of the pair was composed in the same way as the pieces of the learning phase had been. All participants reported extreme difficulties in performing the task. Numerous participants complained that it was difficult to differentiate the two pieces of the pairs. Both experimental groups nevertheless performed above chance with 61% correct responses for non-musicians and 62% of correct responses for musicians, and with no significant difference between the two groups. In a second experiment (run with musically untrained listeners only), the stimuli of the learning phase were identical to those of the previous experiment, whereas the stimuli of the test phase consisted of pairs in which one of the pieces was derived from a retrograde inversion of the tested row. The striking finding was that the participants continued to discriminate grammatical from ungrammatical pieces above chance (60% of correct responses), suggesting that even musically untrained listeners are able to internalise via passive exposure complex regularities derived from the twelve-tone technique. This conclusion is consistent with other findings showing that the structures of Western contemporary music are processed in a similar way by musically trained and untrained listeners. After a short exposition phase, listeners were sensitive to the structure of twelve-tone music. The perception of this music is assumed to be based on frequency distributions of tone intervals. These results shed some light on the implicit versus explicit nature of the acquired knowledge, and the content of the information internalised through listening to these pieces.

Most probably, the knowledge internalised during the listening to the serial pieces was inaccessible to the explicit thought of the participants. If knowledge internalised through exposure was represented at an explicit level, then experts should be more able than non-expert participants to explicitly use this knowledge. This difference should result in a clear advantage for musical experts over musically untrained listeners. If, however, the acquired knowledge is represented at an implicit level, no strong difference should be observed between musically expert and novice participants. The present study converges with conclusions drawn from several other studies run with Western tonal music and argues in favor of the implicit nature of acquired musical knowledge.

Figure 2.3: Higher panel: One of the two twelve-tone row used in the study (the "grammatical" one). The row is shown in these four basic forms: original (O), inverted (INV), retrograde (RET), and retrograde inversion (RET INV). Lower panels: Example of pairs of matched pieces composed using two different rows ("grammatical" and "ungrammatical"). Both pieces share the same superficial features (rhythm, pitch ranges, overall form of melodic contour, duration, dynamics).

2.2 Perspectives in musical learning: using multimedia technologies

2.2.1 How should the learning of Western tonal music be optimised with the help of multimedia technologies?

Explaining the theoretical core of the Western musical system is one of the most difficult tasks for music teachers, and it is generally assumed that this explanation should only occur at the end of the curriculum in both music conservatoire and university departments. Lerdahl's Tonal Pitch Space Theory (TPST, Lerdahl, 2001) is likely to contribute to the development of music tools that would help music lovers as well as those at an early stage of musical study to improve their understanding of Western tonal music. The TPST can be considered as an idealised knowledge representation of tonal hierarchy. The psychological representation of knowledge implies a certain number of questions, for which different solutions have been proposed (Krumhansl et al., 1982a,b; Krumhansl and Kessler, 1982; Longuet-Higgins, 1978). For all these approaches, tonal hierarchies are represented in the form of a multidimensional space, in which the distances of chords from the instantiated tonic correspond to their relative hierarchical importance. The more important the chord is, the smaller the distance. Lerdahl successfully explains the way in which the TPST synthesises various existing musicological and psychological models and suggests new solutions. In the opinion of the authors, the crucial contribution of the model is the description of a formal tool to quantify the tonal distances between any couple of events belonging to any key, a quantification that no other approach proposes.

The TPST model outlines several developments to the model initially described in an earlier series of articles (Lerdahl, 1988, 1991). We summarise here the basic ideas. According to the theory, tonal hierarchy is represented in three embedded levels. The first two (the pitch class level and chordal level) represent within-key hierarchies between tones and chords. The third level represents the distances between keys (region level). The pitch class level (basic space) represents the relation between the 12 pitch classes. It contains five sublevels (from level a to e), corresponding to the chromatic level (level e), diatonic level (level d), triadic level (level c), fifth level (level b) and the tonic level (level a). In a given context, a tonic tone, part of a tonic chord, will be represented at all five levels. The fifth and the third tones of a tonic chord will be represented at four levels (from b to e) and three levels (from c to e) respectively. A diatonic but non-chordal tone will be represented at two levels (from d to e). A non-diatonic chord will be represented at only one level (level e). The level at which a given pitch class is represented thus reflects its importance in the tonal context. For example, in the C major key, the tone C is represented at all levels (from a to e), the tone G, at four levels (from b to e), the tone E, at three levels (from c to e) and the diatonic tones of the C major scale are represented at two levels only (from d to e).

This representation has two implications. First, it allows an understanding as to why tones (e.g. of a C major chord), which are distant in interval (C-E-G-C), can nevertheless be perceived being as close as are adjacent notes (C-D-E-F-G). Though forming distant intervals, the notes of the chord are adjacent at the triadic level in the representational space (level c). Moreover, this explanation of musical tension bound to these forces of attraction constitutes a very promising development for psychology. The second implication concerns the computation of distances between chords. If the C major chord was played in the context of G major, the tone F# will be represented at two levels (from d to e), while the tone F would remain at only one level (level e). This would produce one change in pitch class. The central idea of the TPST is to consider the number of changes that occurs in this basic space when the musical context is changed (as in the present example) as a way to define the pitch-space distance between two musical events.

The second level of the model involves the chordal level, that is the distance between chords in a given key. The model computes the distances separating the seven diatonic chords taking into account the number of steps that separate the roots of the chords along the circle of fifths (C-G-D-A-E-B-F) and the number of changes in pitch-class levels created by the second chord. Let us consider the distance between the C and G major chords in the key of C major. The G major chord induces 4 changes in the pitch-class level. The dominant tone D is now represented at 2 additional levels (from b to e), the third tone B at one additional level (from c to e) and the tonic tone at one additional level (from a to e). The number of steps that separates the two chords on the circle of fifths equals 1. As a consequence the tonal pitch-space distance between these two chords in this key context equals 5. Following the same rationale, the distance in pitch-space between the tonic and the subdominant chords equals 5. The distance between the tonic and the submediant (sixth degree) chords equals 7, as does the distance between the tonic and the mediant chords (third degree). The distance between the tonic chord and the supertonic (second degree) equals 8 as does the distance between the tonic and the diminished seventh chords (seventh degree). This model quantifies the strength of relations in harmonic progressions. Accordingly, the succession tonic/submediant (I/vi) corresponds to a harmonic progression that creates stronger tension than the succession tonic/subdominant (I-IV).

The third level of the TPST model involves the regional level. It evaluates distances between chords of different regions by taking into account the distances between regions as well as the existence of a pivot region. The regional space of the TPST is created by combining the cycle of fifths and the parallel/relative major-minor cycle. That is to say, the shortest distance in regional space (i.e., 7) is found between a given major key (say C major) and its dominant (G), its subdominant (F), its parallel minor (C minor) and its relative minor key (A minor). The greatest distance (30) is found between a major key and the augmented fourth key (C and F#). The tonal distance between two chords of different keys depends on the musical interpretation of the second chord. For example, in the context of C major key the distance between a C major chord and a C# minor chord would equal 23 if the C# is interpreted as a sixth degree (vi) of the E major key. The distance equals 30 if the C# is understood as the tonic chord of the Db minor key. As a consequence, the distance in pitch-space between two events that belong to distant keys depends on the selected route between the two events. In most cases, the selected route is defined by the overall musical context. By default, the model computes this distance according to the principle of the shortest path: "the pitch-space distance between two events is preferably calculated for the smallest value" (Lerdahl, 2001, p.74). The shortest path principle is psychologically plausible. It has the heuristic merit of being able to influence the analysis of time-span and prolongational reductions (Lerdahl and Jackendoff, 1983) by preferring an analysis that reduces the value of these distances. The implementation of this principle in an artificial system should fairly easily lead to "intelligent" systems capable of automatic harmonic analysis.

One of the main features of the TPST as an efficient learning tool is to bridge the intuitive mental representations of untrained listeners with the mental representations of experts. Current developments in multimedia offer considerable opportunities to evolve the naive representation of novices in a given domain. The basic strategy consists in combining different modes of knowledge representation (e.g. sounds, image, language, animation) to progressively transform the initial mental representation into a representation of the domain that fits as closely as possible with that of experts. In the present case, the use of a space to describe the inner structure of the Western tonal system considerably facilitates this transformation. The mental representation of a complex system in a two or three-dimensional space is a metaphor that is common in a large variety of domains and that is intuitively accessible even for a child. A musical learning tool may thus consist of a multimedia animation that illustrates how music progresses through pitch-space. As it is shown in Figure 2.4, the animation displays in real-time every distance travelled through pitch space. After having listened several times to the piece, the journey through the pitch-space of the piece would be stored in memory in both visual and auditory formats. After listening to several pieces of the same stylistic period, the journeys through the pitch-space specific to this style would be stored in memory. After listening to several pieces of the Western music repertoire, listeners would create a mental representation of the overall structure of the tonal pitch space that fits with that of the experts. From a teaching perspective, the interesting point is that this mental representation will emerge from mere exposure to musical pieces presented with this music tool. In other words, the tool allows a passive exploration of the tonal pitch space by visualizing in a comprehensible format the deep harmonic structure of the heard pieces.

The structure of the space can be adapted at will and should notably be adjusted to suit the age of the user. At this early stage of development of the multimedia tool, we chose a structure that mimics real space with planets and satellites. Given the circularity of the Western musical space, only one portion of the space can be seen at a given time point, but this portion will progressively change when the music is moving from one region to another. A planet metaphorically represents a key, while the satellites represent the seven diatonic chords. Satellites corresponding to played chords are lit up in yellow, thus representing the route of the harmonic progressions within each key. The colour of the planet representing the key intensifies when several chords from the key are played, thus imitating the fact that the feeling of the tonality increases with duration. When the music modulates to another key, chords from both the initial key and the new key light up, and the animation turns towards the new key, and then discovering another portion of the tonal pitch-space. When the piece of music progresses rapidly towards distant keys, as in the case of Chopin's *Prelude* in E major, the pivot keys are briefly highlighted and passed quickly. The journey depends upon the modulations that have occurred. With the present tool, the user can associate the visual journey through tonal pitch space with the auditory sensation created by the music. The animation contains sufficient music theoretic information to allow the user to describe this musical journey in terms that are close to those employed by musicologists. Of course, this animation may also bring other important elements for the comprehension of harmonic processes, such as the arrangement of chords and voice leading. Connected to a MIDI instrument, it may equally be transformed into a tool for tonal music composition. By chaining chords together, the user can follow his or her journey through tonal space, and explore the structure of the tonal space.

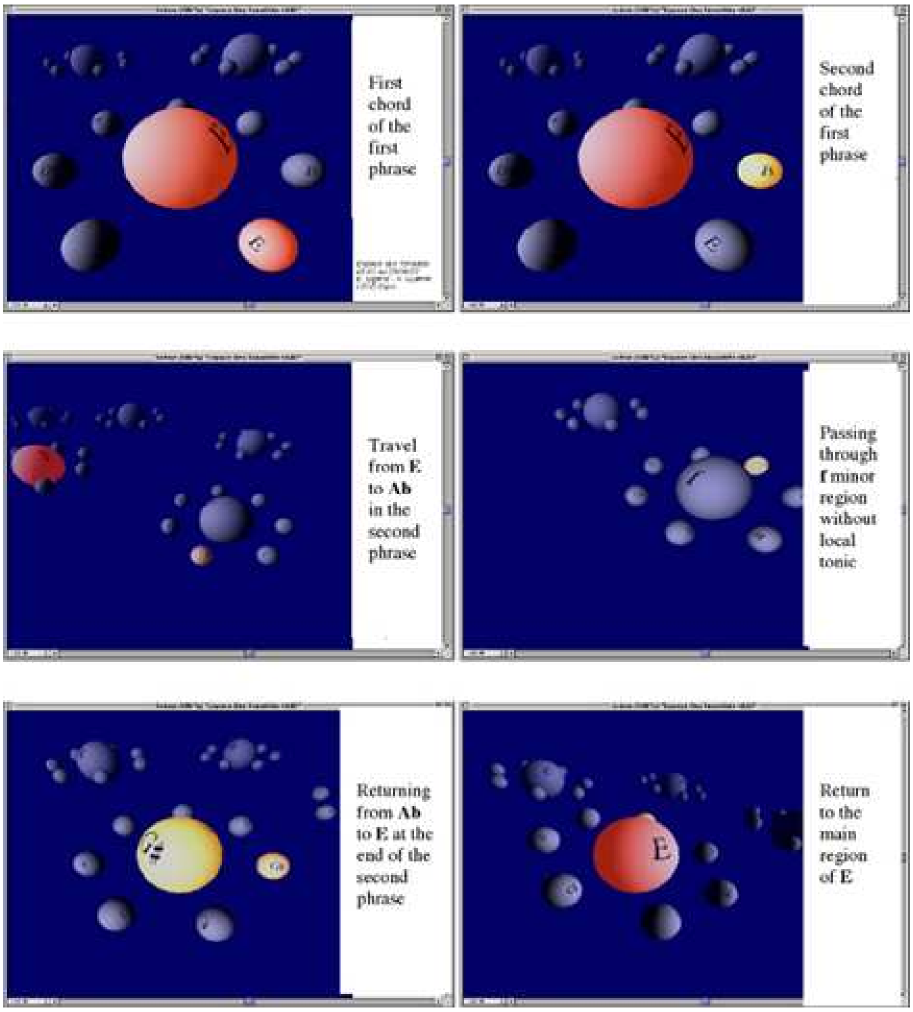

Figure 2.4: Musical tool derived from TPST and which is currently being developed by LEAD - This animation is available at the address `http://www.u-bourgogne.fr/LEAD/people/bigand_e.html`. This research is supported by a CNRS grant “Société de l’information”. The animation was realised in collaboration with Aristide Quenel.

2.2.2 Creating learning multimedia tools for music with the contribution of cognitive sciences and ergonomics

The first multimedia works comprising music began to emerge at the beginning of the nineties. Since then, the number and diversity of musical multimedia products (CD-Rom, DVD-Rom, web sites) have been increasing considerably. However, multimedia products helping the user to integrate musical structures are rare. If we want to propose successful learning multimedia tools, the main question is how to use multimedia resources and why. We are going to present some assumptions of cognitive ergonomics that seem fundamental to multimedia dedicated to music education. These assumptions will be illustrated by two multimedia learning tools created at the LEAD. The perceptual supposed advantage of the tools will be evaluated during a next phase by several tests concerning attentional processing, memory and understanding.

The first principle is that the vantage point of music learning tools should be immediate representation formats of the non-experts. The tools have to combine in an advantageous way the various possibilities of multimodal representations to make these initial representations evolve towards those of experts (the principle of affordance; Gibson, 1977). Multimodality should be used as a powerful means to clarify the structure of complex systems, and to allow the user to easily develop a mental representation of the system. This mental representation should be compatible with the representation of experts. The aim of the project was to give listeners the access to a musical system that is often considered as complex (contemporary music) while being potentially of high educational value. It was an ideal opportunity to try out a multimedia approach to the learning of a complex system.

Reduction of information and optimisation of presentation forms

One of the main problems concerning multimedia is an overload of presentation forms. Multiplication of presentation forms (text, picture, animation, video, sound, etc.) often entails a cognitive cost that is high compared to the benefits in terms of training. This profusion of presentation forms often leads to an explosion of the quantity of information presented to the user, and to the lack of an objective analysis of the presentation forms used and of the combination of learning modes (visual & auditory, verbal & visual, etc.) available in the proposed tools. Information overload is often accompanied by an organisation of knowledge based on models that are not adapted to the initial knowledge of a user. The second fundamental principle in producing multimedia learning tools is thus reducing the quantity of information and optimizing the form in which it is presented. Properly used multimodality, particularly concerning the interaction between vision and audition, improves attentional processes, memorisation of musical material, and develops the capacity to represent the musical structures. There are several types of representation of the musical structures. The score is the best known of the notation. However, it requires specific knowledge and regular musical practice. There are other forms of music representation: tablatures of string instruments, sonograms (time-evolving spectrum), wave forms (amplitude), tracks or piano-rolls of sequencer softwares, etc. All these representation modes can certainly be employed in multimedia, but they often require expert knowledge.

The project carried out at the LEAD used graphical representations that can advantageously replace the representation forms of experts. These graphics consist of simple forms symbolizing one or more elements of musical structure (melodic contour, texture, harmonic density, rhythmic pattern, etc). The principal constraint is that these forms should not require additional coding, but induce the musical structure in an intuitive and direct way. Other presentation forms, which require expert knowledge, never intervene in the initial presentation of a musical excerpt. Figures 2.5 and 2.6 show two representation forms of a chord sequence in the piece *Couleurs de la Cité Céleste* by Oliver Messiaen. The constitution in terms of tones is identical for the 13 chords. It is the register, the duration and the change in instrumentation between the various instruments which give listeners an impression of a succession of sound colours (*Klangfarbenmelodie*). The excerpt is represented by blocks of colours whose width corresponds to the duration of the chords. The choice of the colours has been determined by the name of the colours written by the composer in the score. Their height symbolises the extent of chords (from the lowest to the highest). The position on the scale (on the left) represents the register. Blocks appear synchronously with the sound. With this type of representation, it is easy, for non-expert listeners, to perceive a degree of similarity between certain chords. Based on this type of representation, a user may intuitively become aware of the external structure of a sequence, even if it is not sufficient to form a precise representation of the musical structure. Figure 2.6 represents the same sequence of chords in form of a score. In order to focus listener's attention on the harmonic structure, the real duration of the chords was replaced by an equal duration for all chords. However, in contrast to a graphical representation of sound that was privileged here, this mode of representation is to give the users an opportunity to decompose musical structures. The users can choose what they want to listen to: the whole sequence, each chord separately, groups of instruments within a chord or each note of a chord.

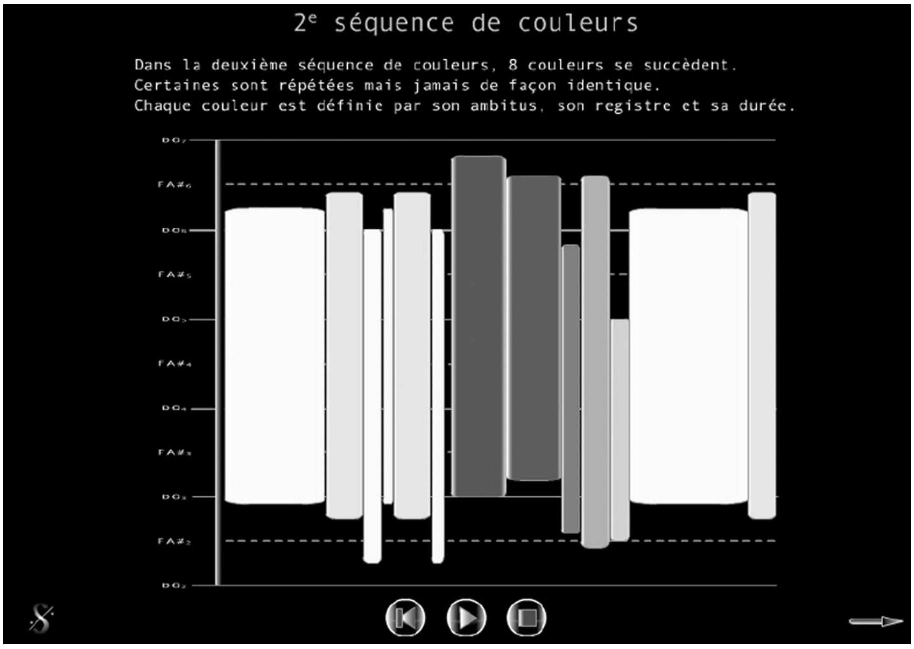

Figure 2.5: Multimedia tool for learning contemporary music. Graphic representation of a chord sequence of *Couleurs de la Cité céleste* by Olivier Messiaen.

Synthesis of knowledge and implementation of continuity

Multimedia tools of learning should synthesise the knowledge of music in order to make it available to non-experts. Thus, it is necessary to implement this knowledge in a way adapted to the initial knowledge of the user. In the case of music (complex music in particular), it is important to raise the question of perceptibility of musical structures. It is a question of knowing exactly what should be emphasised. In our project, the pieces were selected according to the cognitive problems they represent (relating to their aesthetic differences). For example, *Couleurs de la Cité Céleste* by Messiaen is representative of the aesthetics where colour and timbre are important. This piece is composed of a great variety of musical elements that follow one another to form a sound mosaic. Globally, a multimedia learning tool must favour categorisation and memorisation of musical material, in order to allow the emergence of the mental representation of a temporal organisation of a piece. Figure 2.7 shows the main page of the multimedia learning tool of Messiaen's piece. One can see the representation of the formal structure of the excerpt in the center of the screen. Eight icons on the right and on the left of the screen give access to eight links (history of the piece, composer's biography, the orchestra, a large-scale structure of

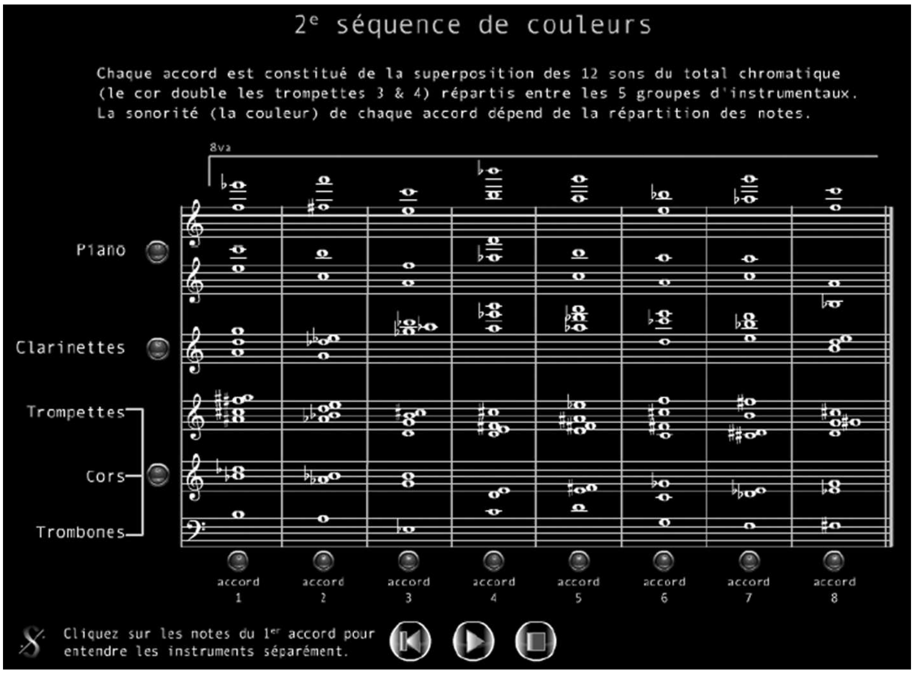

Figure 2.6: Multimedia tool for learning contemporary music. Score representation of the same chord sequence as Figure 2.5.

the excerpt, and four of the main materials of the piece: texts and metaphors of the Apocalypse, Gregorian chant, colours, and bird songs).

One of the crucial problems in the pedagogy of listening is that of attention. Perception of musical structures is strongly dependent on attentional processes. In the simplest case, attentional processes are guided by the music itself (when, for example, a composer emphasises the principal melody by a discrete accompaniment) or by the performer (when he chooses to emphasise a specific structural element). However, most of the time, music has a complex and deliberately ambiguous structure. Contrary to the traditional methods in music education, multimedia tools make it possible to easily focus the listener's attention on internal or external elements of musical structure.

One part of our project consisted in seeking in the resources of multimedia the means of guiding attentional processes and, beyond, of favouring the memorisation and the comprehension of musical structures. The schematic representation of the formal structure of the beginning of *Couleurs de la Cité Céleste* (Figure 2.7) was conceived to facilitate mental representation of a

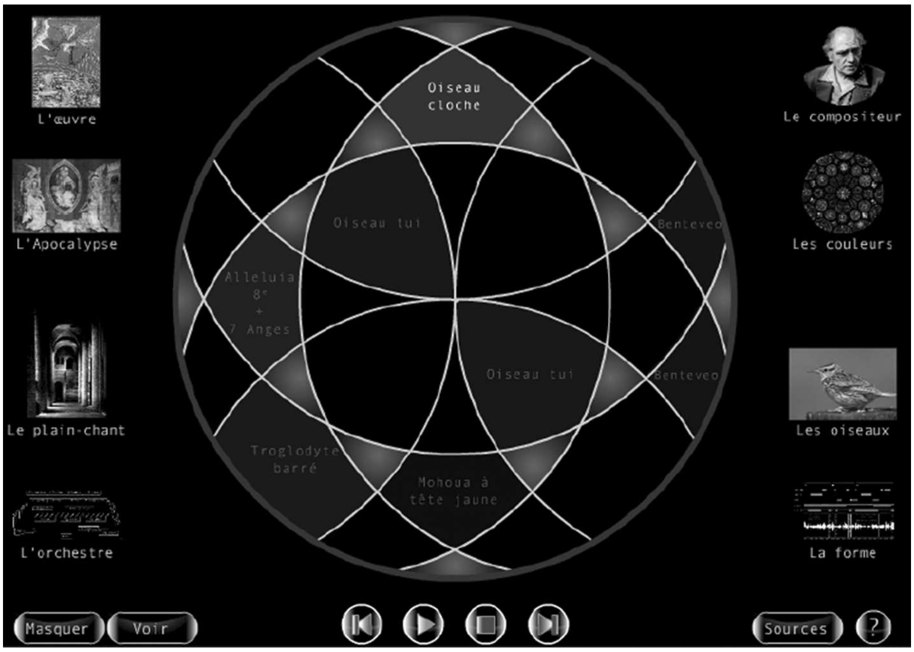

Figure 2.7: Multimedia tool for learning contemporary music. Main page of the multimedia learning tool of *Couleurs de la Cité céleste* de la Cité céleste by Olivier Messiaen.

complex formal structure in which short musical sequences follow one another in form of a mosaic that would emerge during perception. This multimedia animation does not contain any oral or textual explanation. Awareness of the structure emerges solely from the visual and auditory interaction. The choice of a circular representation corresponds to the form of a piece whose elements return in a recurrent and circular way. Each piece of the mosaic corresponds to a short musical sequence. Each colour represents a type of musical material (e.g. blue for bird songs). Nuances of colours differentiate the variations inside each category of sequence. The animation takes into account the cognitive processes of attention and memorisation. At the beginning of the animation, the stained-glass scheme is empty. Progressively, as the music unfolds, empty spaces are filled until the stained-glass is completed (all the sequences were played). When a sequence has ended, the corresponding stained-glass is gradually obscured (approximately 6 to 7 seconds, according to the maximum duration of the perceptual present; Fraisse, 1957). The luminosity of the piece of stained-glass solidifies at a very low rate. This process is very close to a trace of an event remaining in memory. When an identical sequence returns, the part that had been previously activated is briefly reactivated and then turns over in stand-by. This multimedia

Figure 2.8: Multimedia tool for learning contemporary music. Formal structure represented in the multimedia learning tool of *Eight Lines* by S. Reich.

artifice supports the categorisation and the memorisation of materials. It also makes it possible to establish bonds of similarity and consequently gives direction to the formal structure which proceeds under the eyes of the user. This example illustrates how it is possible to guide listeners to focus their attention on sequential events. It is also useful to focus the attention of the user on simultaneous events. Figure 2.8 shows the animated representation of the formal structure of the beginning of *Eight Lines* by S. Reich. In contrast to the Messiaen's piece represented by stained-glass, the piece by Reich, whose unfolding follows a linear trajectory, is represented by 8 rectangular boxes. Coloured rectangular paving stones indicate the moment of appearance and the duration of intervention of each instrumental part. Synchronisation between the sound and the image is visualised by a vertical marker. When an instrument is active, its coloured paving stone is cleared up. In Figure 2.8, the active instrumental parts are the parts of viola and violoncello (lines 7 and 8, at the bottom), flute and bass clarinet (lines 1 and 2, at the top). Inside a paving stone, graphical animations, always synchronised with music, emerge to focus the attention on the melody of the instrumental part. The points indicate the onset of pitches, the lines indicate the melody contour.

New technologies of sound processing may provide other new possibilities for multimedia learning tools for music. The sound files obtained with these techniques or especially dedicated software can be integrated into the multimedia in order to improve the learning tools. The possibility of an interactive deconstruction or reconstruction of the musical structures combined to specific visual interfaces is certainly the most promising perspective in the nearest future.

Conclusion

The power of implicit learning is without doubt one of the major contributions of research into musical cognition. It reduces the distance between listeners (non-musicians and musicians) and leads to question the common practices in music education. Implicit learning supplies a solid scientific basis, together with the contributions of cognitive psychology (memory and attending processes), ergonomics and new technologies to create multimedia learning tools for music.

Bibliography

G. T. Altmann, Z. Dienes, and A. Goode. Modality independence of implicitly learned grammatical knowledge. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 21:899–912, 1995.

M. Ayari and S. McAdams. Aural analysis of arabic improvised instrumental music (tagsim). *Music Perception*, 21:159–216, 2003.

J. J. Bharucha. Music cognition and perceptual facilitation: A connectionist framework. *Music Perception*, 5(1):1–30, 1987.

E. Bigand and B. Poulin-Charronnat. Are we "experienced listeners"? a review of the musical capacities that do not depend on formal musical training. *Cognition*, 100(1):100–130, 2006.

E. Bigand, P. Perruchet, and M. Boyer. Implicit learning of an artificial grammar of musical timbres. *Cahiers de Psychologie Cognitive/Current Psychology of Cognition*, 17(3):577–600, 1998.

E. Bigand, F. Madurell, B. Tillmann, and M. Pineau. Effect of global structure and temporal organization on chord processing. *Journal of Experimental Psychology: Human Perception and Performance*, 25(1):184–197, 1999.

H. Budge. *A study of chord frequencies*. Bureau of Publications, Teachers College, Columbia University, New York, 1943.

C. Delannoy. Detection and discrimination of dodecaphonic series. *Interface*, 1:13–27, 1972.

Z. Dienes and C. Longuet-Higgins. Can musical transformations be implicitly learned? *Cognitive Science*, 28:531–558, 2004.

Z. Dienes, D. Broadbent, and D. C. Berry. Implicit and explicit knowledge bases in artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 17: 875–887, 1991.

J. W. Dowling. Recognition of melodic transformations: inversion, retrograde, retrograde inversion. *Perception and Psychophysics*, 12:417–421, 1972.

J. W. Dowling and D. L. Harwood. *Music Cognition*. Academic Press, Orlando, Florida, 1986.

P. Fraisse. *Psychologie du temps*. Presses Universitaires de France, Paris, 1957.

R. Francès. *La perception de la musique*. Vrin, Paris, 1958.

J. J. Gibson. The theory of affordances. In R. E. Shaw and J. Bransford, editors, *Perceiving, Acting, and Knowing*. Lawrence Erlbaum Associates, Hillsdale, 1977.

P. Griffiths. *A Concise History of Modern Music*. Thames and Hudson, London, 1978.

S. Grossberg. Some networks that can learn, remember and reproduce any number of complicated space-time patterns. *Studies in Applied Mathematics*, 49:135–166, 1970.

S. Grossberg. Adaptive pattern classification and universal recoding: I. parallel development and coding of neural feature detectors. *Biological Cybernetics*, 23:121–134, 1976.

J. H. J. Howard and J. A. Ballas. Syntactic and semantic factors in the classification of nonspeech transient patterns. *Perception and Psychophysics*, 28(5):431–439, 1980.

J. H. J. Howard and J. A. Ballas. Acquisition of acoustic pattern categories by exemplar observation. *Organization, Behavior and Human Performance*, 30:157–173, 1982.

E. K. Johnson and P. W. Jusczyk. Word segmentation by 8-month-olds: When speech cues count more than statistics. *Journal of Memory and Language*, 44(4):548–567, 2001.

T. Kohonen. *Self-Organizing Maps*. Springer, Berlin, 1995.

C. L. Krumhansl. *Cognitive foundations of musical pitch*. Oxford University Press, New York, 1990.

C. L. Krumhansl and E. J. Kessler. Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys. *Psychol Rev*, 89(4):334–368, 1982.

C. L. Krumhansl and G. Sandell. The perception of tone hierarchies and mirror forms in twelve-tone serial music. *Music Perception*, 5(1):31–78, 1987.

C. L. Krumhansl, J. Bharucha, and M. A. Castellano. Key distance effects on perceived harmonic structure in music. *Percept Psychophys*, 32(2):96–108, 1982a.

C. L. Krumhansl, J. J. Bharucha, and E. J. Kessler. Perceived harmonic structures of chords in three related keys. *Journal of Experimental Psychology: Human Perception and Performance*, 8: 24–36, 1982b.

F. Lerdahl. Tonal pitch space. *Music Perception*, 5(3):315–349, 1988.

F. Lerdahl. Structure de prolongation dans l'atonalite. In S. McAdams and I. Deliege, editors, *La musique et les sciences cognitives*, pages 171–179. Mardaga, Liège, 1989.

F. Lerdahl. Pitch-space journeys in two chopin preludes. In M.R. Jones and S. Holleran, editors, *Cognitive Bases of. Musical Communication*. Amer Psychological Assn, Washington, 1991.

F. Lerdahl. *Tonal Pitch Space*. Oxford University Press, Oxford, 2001.

F. Lerdahl and R. Jackendoff. *A Generative Theory of Tonal Music*. MIT Press, Cambridge (MA), 1983.

H. Longuet-Higgins. The perception of music. *Interdisciplinary Science Review*, 3:148–156, 1978.

F. Meyer. *Emotion and Meaning in Music*. University of Chicago Press, Chicago and London, 1956.

E. Narmour. *The Analysis and Cognition of Basic Melodic Structures*. University of Chicago Press, Chicago, 1990.

A. S. Reber. Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior*, 6:855–863, 1967.

A. S. Reber. Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General*, 118:219–235, 1989.

D. E. Rumelhart and D. Zipser. Feature discovery by competitive learning. *Cognitive Science*, 9: 75–112, 1985.

J. R. Saffran, E. L. Newport, and R. N. Aslin. Word segmentation : The role of distributional cues. *Journal of Memory and Language*, 35(4):606–621, 1996.

C. A. Seger. Implicit learning. *Psychological Bulletin*, 115:163–169, 1994.

E. D. Thiessen and J. R. Saffran. When cues collide: use of stress and statistical cues to word boundaries by 7- to 9-month-old infants. *Developmental Psychology*, 39(4):706–716, 2003.

B. Tillmann and S. McAdams. Implicit learning of musical timbre sequences : statistical regularities confronted with acoustical (dis)similarities. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 30:1131–1142, 2004.

B. Tillmann, J. J. Bharucha, and E. Bigand. Implicit learning of tonality: a self-organizing approach. *Psychological Review*, 107(4):885–913, 2000.

B. Tillmann, J. J. Bharucha, and E. Bigand. Implicit learning of regularities in western tonal music by self-organization. In R. French and J. Sougné, editors, *Perspectives in Neural Computing series*, pages 175–184. Springer Verlag, London, 2001.

C. von der Malsberg. Self-organizing of orientation sensitive cells in the striate cortex. kybernetic. *Biological Cybernetics*, 14:85–100, 1973.